Proceedings – 23rd Annual Conference – IEEE/EMBS Oct.25-28, 2001, Istanbul, TURKEY

AUTOMATED SLEEP STAGE SCORING BY DECISION TREE LEARNING

Masaaki Hanaoka, Masaki Kobayashi, Haruaki Yamazaki
Faculty of Engineering,Yamanashi University
4-3-11 Takeda, Kofu, Yamanashi, Japan

Abstract- **In this paper we describe a waveform recognition method that extracts characteristic parameters from waveforms and a method of automated sleep stage scoring using decision tree learning that is in practice regarded as one of the most successful machine learning methods. In our method, first characteristics of EEG, EOG and EMG are compared with characteristic features of alpha waves, delta waves, sleep spindles, K-complexes and REMs. Then, several parameters that are necessary for sleep stage scoring are extracted. We transform these extracted parameters into a few discrete variables using canonical discriminant analysis and the discretization method based on a random walk, and then a committee that consists of several small decision trees is formed from a small number of training instances. Furthermore final sleep stages are decided by a majority decision of the committee. Our method was applied to the digitized PSG chart data, provided by the Japan Society of Sleep Research and we carried out an evaluation experiment. The experiment indicated that our method can quickly execute learning and classification and precisely score sleep stages.**

Keywords- **EEG, sleep stages, waveform recognition, decision tree learning**

I. INTRODUCTION

In general, characteristics of a waveform can be expressed quantitatively by using methods such as FFT, wavelet transform, digital filters, which extract signals of specific frequencies, and averaging, so as to eliminate noises. As a rule, a bio-signal changes in a complicated manner and does not necessarily correspond to changes in a living organism. Hence, it has been difficult to obtain useful information on the interpretation of bio-signals with these quantitative methods alone. For this reason, various waveform recognition methods have been proposed with computer processing replacing visual inspection by specialists. Furthermore many studies have been carried out on automated sleep stage scoring based on quantified features of bio-signals. In fact, some methods have been proposed [1] [2] [3] [4] that exceed the agreement percentage among specialists (approximately 70%).

In this paper, we propose a new method of waveform recognition and an automated sleep stage scoring by decision tree learning.

II. EXTRACTION OF CHARACTERISTIC PARAMETERS

To achieve an automatic inspection of polysomnogram (PSG) based on R&K rules, it is necessary to detect characteristic waves from EEG, EOG and EMG at every epoch (a time unit for sleep stage scoring, usually 20 or 30 sec).

A. Waveform Recognition of EEG

In our method the waveform recognition of characteristic waves (alpha waves, delta waves, sleep spindles and K-complexes) is carried out based on directions, peaks, bottoms, and durations of waves which appear on EEG.

In the waveform recognition, 150 sample data points at the beginning of the epoch are collected and the average is calculated to determine a baseline. Next, prior to peak detection, a positive detection level and a setting level are predetermined for each characteristic wave. The positive detection level is assigned above the baseline. A point is a peak if it is the maximum value in a control interval (Fig.1). The control interval is defined as the interval from where the waveform first crosses the positive detection level to where it is smaller by the amount of the setting level from the maximum value. Similarly, a bottom (negative peak) is detected. At every control interval, a peak and a bottom are alternately detected.

Next, the period of *"the wave"* as indicated by the time interval between two adjacent bottoms and the average amplitudes from the bottom to the peak in the first half of *the wave* and the second half of *the wave* are calculated (Fig.2). We then determine whether the two quantities are inside of the region of frequency and amplitude based on the properties of characteristic waves. If both the frequency and the amplitude are inside of the region, *the wave* is recognized as a characteristic wave. Otherwise *the wave* is ignored.

From the duration and the total appearance time of *the wave* detected by the method described above, the characteristic parameters of EEG are obtained.

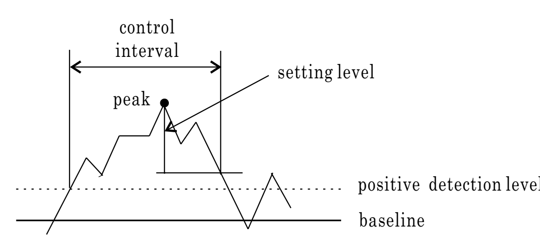

Fig. 1. Detection of EEG peak.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle Automated Sleep Stage Scoring by Decision Tree Learning	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Yamanashi University Faculty of Engineering 4-3-11 Takeda, Kofu, Yamanashi, Japan	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Oct 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

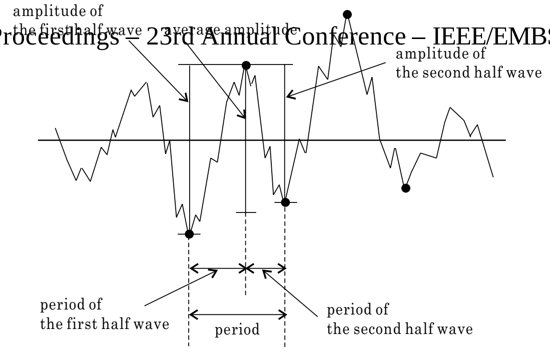

Fig. 2. The quantified features of EEG.

B. Detection of REMs

Two EOG signals that are induced by electrodes attached outside of the eyes during REM sleep are observed as two rectangular waves with counter-phases. Our detection algorithm of REMs works as follows:

1. The onset level and the offset level of REMs are determined from the onset angle (40°) and the offset angle (28°), respectively. And a pointer is set at the beginning of the epoch.
2. To detect the onset of REMs, the indicator is calculated at the location indicated by the pointer.
3. If the indicator exceeds the onset level, the location is chosen as the onset of REMs and the pointer is moved 12 points ahead. Otherwise the pointer is moved 1 point ahead and the operation returns to step 2.
4. To detect the offset of REMs, the indicator is calculated at the location indicated by the pointer.
5. If the indicator is less than the offset level, the location is chosen as the offset of REMs and the pointer is moved 12 points ahead, and then returns to step 2. Otherwise the pointer is moved 1 point ahead and the operation returns to step 4.

The algorithm employs the product of two differences as the indicator. One is the difference in the amplitudes between the data indicated by the pointer and the data 6 points ahead. Another is the difference between the data 6 points ahead and 12 points ahead.

When the left-eye wave is recognized as the inverse of the right-eye wave, we employ the number of such locations as one of characteristic parameters.

C. Detection of EMG

EMG shows an extremely large amount of electric discharge on the occasion of stage wake (SW) or movement time (MT). As a sleep deepens from stage 1 (S1) to stage 4 (S4), EMG goes lower in amplitude, but continues to discharge and completely disappears at stage REM (SREM). For sleep stage scoring, it is necessary to examine the amount of electric discharge.

In this study, a total electric discharge of EMG in the epoch, namely the integrated values of the absolute values of EMG amplitude is normalized by the electric discharge of the setting REM level (5 mV×the number of sampling data in 1 epoch). We employ it as one of characteristic parameters.

III. DECISION TREE LEARNING

In order to make a highly accurate automated sleep stage scoring, we use decision tree learning.

The training data in decision tree leaning consists of the explanatory attributes and a target attribute. The learning algorithm generates the tree to classify instances according to the target attribute values (classes).

The points to take into consideration in decision tree learning are as follows:

1. Attribute selection: For effective classification, the appropriate explanatory attributes must be selected.
2. Pruning: To prevent the over-fitting for the training data, the decision tree must be pruned.
3. Discretization of continuous attributes: For effective classification, some continuous valued explanatory attributes must be discretized.

In our study we employed C4.5 [5] for attribute selection and pruning. For sleep stage scoring, it is necessary to discretize a large amount of continuous attribute data. Our discretization method RWS [6] is suitable for this purpose and is used in this system. Furthermore, canonical discriminant analysis [7] and committee learning by bagging [8] are employed in order to make a compact decision tree.

Fig.3 shows the processing flow of decision tree learning for automated sleep stage scoring.

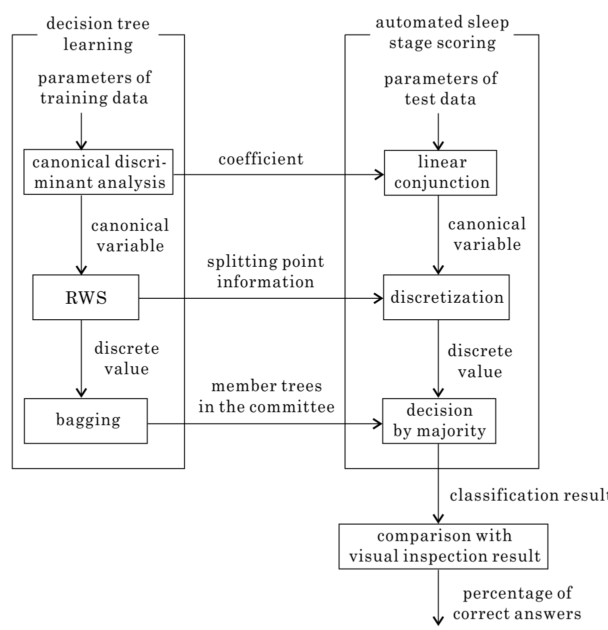

Fig. 3. Flowchart of processing of decision tree learning and automated sleep stage scoring.

Table 1. The values of characteristic parameters extracted from the digitized PSG chart data.

sleep stages \ parameters	appearance ratio of alpha waves	appearance ratio of delta waves	appearance ratio of sleep spindles	total numbers of K-complexes	numbers of REMs	normalized integrated EMG
SW	35.24 ± 12.38	0.00 ± 0.00	8.21 ± 6.69	0.00 ± 0.00	0.53 ± 0.58	12.81 ± 8.82
SREM	11.28 ± 9.76	0.05 ± 0.79	1.14 ± 1.96	0.00 ± 0.00	0.65 ± 0.68	1.18 ± 0.95
S1	6.18 ± 7.05	0.19 ± 1.35	1.10 ± 1.99	0.01 ± 0.07	0.00 ± 0.00	4.76 ± 4.25
S2	6.01 ± 5.63	4.37 ± 4.43	6.79 ± 4.34	0.02 ± 0.13	0.00 ± 0.03	2.34 ± 1.86
S3	1.76 ± 2.41	34.64 ± 8.39	1.71 ± 1.27	0.11 ± 0.33	0.05 ± 0.21	2.22 ± 2.14
S4	0.74 ± 1.43	42.51 ± 8.34	0.83 ± 1.10	0.15 ± 0.38	0.08 ± 0.28	1.70 ± 0.29
MT	1.11 ± 4.59	0.40 ± 1.99	0.11 ± 0.53	0.20 ± 0.41	0.00 ± 0.00	19.04 ± 8.55

The values on the left are the average values, and the values on the right are the standard deviations.

IV. EVALUATION EXPERIMENT

A. *Experimental Method*

In the evaluation experiment, we used the digitized PSG chart data (the PSG digital data) provided by the Japan Society of Sleep Research (JSSR). In the data, an all-night sleep PSG of 8 hours (288000 seconds) by a 28-year old normal male was recorded in digital form with a 500 Hz sampling rate and 16-bit quantified rate. The sleep stages of 1440 epochs (20 sec/epoch) are provided as the classification results of specialists' visual inspection. The sleep stages were scored by consultation with JSSR computer committee members.

To test the effectiveness of our waveform recognition based method and the decision tree learning, we used 5 kinds of digital bio-signals – 2 channels of EEG (C3, O1), 2 channels of EOG (right eye, left eye), 1 channel of EMG (mentalis muscle) – and compared the classification results using our method to the results obtained by JSSR.

First, the PSG digital data was split into 1440 epochs, 20 sec for 1 epoch. Our waveform recognition was applied to 5 kinds of bio-signals at each epoch, and 6 kinds of characteristic parameters needed for sleep stage scoring were extracted.

Next, the cross-validation was carried out on the data set with 1440 instances which include these 6 kinds of parameters as explanatory attributes and the scoring results by JSSR as a target attribute. In the cross-validation, the data set was randomly divided into five subsets with equal numbers of instances. Then one subset was used as test data and the others as training data. The decision tree learning and the automated sleep stage scoring were applied to all subsets 5 times. The classification accuracy of our method is defined as the average percentage of correct answers over 5-time executions. The learning time and the classification time of our method are defined as the average times for the decision tree learning and for automatic sleep stage scoring respectively.

In this experiment, for pruning, the confidence level of the binomial distribution was set to 25%. For RWS, the significance level was set to 5%. For the committee learning, the sampling ratio of the bootstrap method was set to 70% and the number of member classifiers was set to 7.

B. *Results of Experiment*

Table 1 shows the values of 6 kinds of characteristic parameters extracted from the PSG digital data using the method in Sec.II. The appearance ratio of alpha waves is the highest at SW, next is SREM, and then S1, S2, S3, MT and S4 in descending order. The ratio has a tendency to decrease from S1 to S4 as sleep becomes more deep. The appearance ratio of delta waves, on the other hand, increases and it exceeds 30% at S3 and S4. The appearance ratio of sleep spindles at SW and S2 is from 4 to 10 times larger than that of other sleep stages, except for MT. The total number of K-complexes and the number of REMs do not exceed 1.0 at any stage, but the total number of K-complexes at S3, S4 and MT and the number of REMs at SREM and SW show larger values than those of other stages. The normalized integration value of EMG is largest at MT with SW, S1, S2, S3, S4 and SREM in descending order. This tendency is the same as for alpha waves. These results show that our method can extract the characteristic parameters, except for K-complexes, approximately the same as does R&K rules.

Table 2 shows the results of the automated sleep stage scoring using the decision tree learning of Sec.III. The percentages of correct answers exceed 80% in 4 out of 5 data sets, and the average for all data sets is 81.4%. Both the average percentage and total percentage are, in descending order, S2, SREM, S3, S1, MT, SW and S4. The higher the percentage of correct answers, the larger the number of instances, except for S3 and S1 where that order is reversed. Particularly at S2 where the number of instances occupies 56.9% of the total, the percentages of correct answers are very high, from 88.5% to 95.9%. On the other hand, in MT, SW and S4 stages, where the number of instances occupies only 3.8% of the total, the percentages are smaller than the average, from 40.0% to 53.0%. This is because the situation where the instances in leaves such as MT, SW, and S4, which contain a small number of instances, are included in large leaves because of the pruning, and are buried as errors.

When learning and classification are carried out 5 times by the cross-validation, the average learning time is 39.54 sec and the average classification time is 3.23 sec. Learning time means the time required for decision tree learn-

Table 2. The percentages of correct answers in the automated sleep stage scoring.

	test data sets 1	2	3	4	5	total	percentages of correct answers
SW	3/4	2/4	0/3	1/2	3/4	9/17	
	75.0	50.0	0.0	50.0	75.0	52.9	50.0 ± 30.6
SREM	29/36	40/51	22/33	33/48	39/47	163/215	
	80.6	78.4	66.7	68.8	83.0	75.8	75.5 ± 7.3
S1	23/36	25/41	19/34	26/45	25/40	118/196	
	63.9	61.0	55.9	57.8	62.5	60.2	60.2 ± 3.3
S2	158/177	140/152	172/186	139/157	141/147	750/819	
	89.3	92.1	92.5	88.5	95.9	91.6	91.7 ± 2.9
S3	18/27	24/31	16/24	25/30	30/43	113/155	
	66.7	77.4	66.7	83.3	69.8	72.9	72.8 ± 7.3
S4	2/4	1/3	2/3	0/1	1/2	6/13	
	50.0	33.3	66.7	0.0	50.0	46.2	40.0 ± 25.3
MT	3/4	3/6	2/5	3/5	2/5	13/25	
	75.0	50.0	40.0	60.0	40.0	52.0	53.0 ± 14.8
total	236/288	235/288	233/288	227/288	241/288	1172/1440	
	81.9	81.6	80.9	78.8	83.7	81.4	81.4 ± 1.8

The upper values are the number of the agreement instances with visual inspection / the number of instances in each sleep stage, and the lower values are the percentages of correct answers.

ing as shown in Fig.3 and classification time means the time required for automated sleep stage scoring as shown in Fig.3. We can see that our method requires shorter processing time and is more efficient than the existing methods [3] [4]. The reasons are: 1) our method needs a certain calculation in the pre-learning procedure such as the generation of composed variables and discretization of these variables, but in the learning process it requires only splitting of the data set into subsets, 2) in the classification process it requires only simple operation and comparison of numerical values at the time of branching, so that it can classify instances without a large amount of calculation.

V. CONCLUSIONS

In this paper, we described the waveform recognition method and the automated sleep stage scoring based on decision tree learning. In the former the characteristic parameters are extracted from waveforms and in the latter the decision by majority is carried out in the committee. To test the effectiveness of our method, the value of the characteristic parameters at every sleep stage, the classification accuracy and the processing time (learning time and scoring time) were observed.

We have also performed an evaluation experiment using the PSG digital data provided by JSSR. In the waveform recognition, the characteristic parameters were extracted in a form similar to the description for R&K rules. Then in the decision tree learning, the sleep stage scoring was made quickly with over 70% correct answers, which is in agreement with the percentage obtained by specialists.

Thus, we conclude that our method is promising for automated sleep stage scoring, and it can achieve learning and classification with high accuracy and quick execution.

REFERENCES

[1] J.C. Principe, S.K. Gala, and T.G. Chang, "Sleep staging automaton based on the theory of evidence," *IEEE Trans. on Biomedical Engineering*, vol.36, no.5, pp.503-509, 1989.

[2] M.T. Hellyar, E.C. Ifeachor, D.J. Mapps, E.M. Allen, and N.R. Hudson, "Expert system approach to electroencephalogram signal processing," *Knowledge-Based Systems*, vol.8, no.4, pp.164-173, 1995.

[3] T. Shimada, T. Shiina, and Y. Saito, "Sleep stage diagnosis considering the context of stage sequences by neural networks," *Trans. IEE Japan*, vol.115-C, no.12, pp.1452-1459, 1995.

[4] K. Inoue, K. Kumamaru, and S. Matsuoka, "Automatic determination algorithm of human EEG sleep stages," *Japanese Journal of Electroencepharography and Electromyography*, vol.25, no.6, pp.427-431, 1997.

[5] J.R. Quinlan, *C4.5: Programs For Machine Learning*, Morgan Kaufman, San Mateo, CA, 1993.

[6] M. Hanaoka, M. Kobayashi, and H. Yamazaki, "RWS (Random Walk Splitting): A random walk based discretization of continuous attributes," *Proc. 6th Pacific Rim International Conf. on Artificial Intelligence*, pp.146-155, 2000.

[7] Y. Tanaka, T. Tarumi, and K. Wakimoto, *Handbook for Statistical Data Analysis with Personal Computers II*, Kyoritsu Shuppan, Tokyo, 1984 (in Japanese).

[8] L. Breiman, "Bagging predictors," *Machine Learning*, vol.24, pp.123-140, 1996.